Generating Artificial Social Networks with Small World and Scale Free Properties

Faraz Zaidi, Arnaud Sallaberry, Guy Melançon

Generating Artificial Social Networks with Small World and Scale Free Properties

Faraz Zaidi, Arnaud Sallaberry, Guy Melançon

RESEARCH REPORT
N° 7861
January 2012
Project-Team GRAVITE

ISSN 0249-6399   ISRN INRIA/RR--7861--FR+ENG

Generating Artificial Social Networks with Small World and Scale Free Properties

Faraz Zaidi*†, Arnaud Sallaberry‡†, Guy Melançon§†

Project-Team GRAVITE

Research Report n° 7861 — January 2012 — 31 pages

Abstract: Recent interest in complex networks has catalysed the development of numerous models to help artificially generate and understand these networks. Watts and Strogatz presented a model [37] to explain how the two properties of small world networks, high clustering coefficient and low average path length appear in networks. Barabási and Albert gave a model [1] to explain how networks with power-law degree distribution arise in networks. From these two ground breaking results, many researchers have introduced different models to explain the appearance of networks with small world and scale free properties in the real world.
In this paper, we focus on social networks and comparatively study the structure of real world and artificially generated networks. The differences and similarities of different models are highlighted and their shortcomings are identified. Further more, we present a new model which produces networks with both small world and scale free properties which are structurally more similar to real world social networks.

Key-words: Complex Networks, Social Networks, Network Generation Models, Scale Free Networks, Small World Networks

* Faraz Zaidi is with PAF Karachi Institute of Economics & Technology, email: faraz@pafkiet.edu.pk

† This was partly done when all authors were simultaneously part of the GRAVITE team.

‡ Arnaud Sallaberry is member of the VIDI (Kwan Li MA) at UC Davis, CA, USA, email: asallaberry@ucdavis.edu

§ Part of this paper was written when G. M. was a visiting scholar at TES Montréal, Canada (Michael McGuffin).

**RESEARCH CENTRE
BORDEAUX – SUD-OUEST**

351, Cours de la Libération
Bâtiment A 29
33405 Talence Cedex

Generating Artificial Social Networks with Small World and Scale Free Properties

Résumé : Recent interest in complex networks has catalysed the development of numerous models to help artificially generate and understand these networks. Watts and Strogatz presented a model [37] to explain how the two properties of small world networks, high clustering coefficient and low average path length appear in networks. Barabási and Albert gave a model [1] to explain how networks with power-law degree distribution arise in networks. From these two ground breaking results, many researchers have introduced different models to explain the appearance of networks with small world and scale free properties in the real world.

In this paper, we focus on social networks and comparatively study the structure of real world and artificially generated networks. The differences and similarities of different models are highlighted and their shortcomings are identified. Further more, we present a new model which produces networks with both small world and scale free properties which are structurally more similar to real world social networks.

Mots-clés : Complex Networks, Social Networks, Network Generation Models, Scale Free Networks, Small World Networks

1 Introduction

A social network can be defined as a set of people, or groups of people interacting with each other [30, 35]. Graphically, a network is represented using a set of *nodes* and *edges*, where *nodes* represent people and *edges* represent interaction between people. An important aspect in the study of social networks is how individuals interact to each other to form large and connected social networks.

Many researchers have studied different structural properties of these networks. Two important and revolutionary results were obtained by the discovery of Small World [37] and Scale free [1] properties of networks. A small world network is a network, when compared with a random graph of same node-edge density, has higher clustering coefficient and the typical distance between any two nodes scales as the logarithm of the number of nodes. The most popular manifestation of the concept of low average path length in social networks is the 'six degrees of separation', uncovered by the social psychologist Stanley Milgram, who concluded that there was a path of acquaintances with a typical length of about six between most pairs of people in the United States [22, 32]. Another important characteristic of these networks is the average clustering coefficient of nodes [37], sometimes referred as Transitivity [24]. The concept is very well known in social networks and can be described as the friend of your friend is likely to be your friend. A scale free network [1] is a network in which a few nodes have a very high number of connections (degree) and lots of nodes are connected to a few nodes. These networks have no characteristic scales for the degrees, hence they are called scale free networks [27].

Since the introduction of these two classes of networks, many researchers have developed different models to explain the appearance of networks with small world and scale free properties in the real world. Social networks also exhibit these two properties at the same time. These models are developed to create a profound understanding of real world networks. The goal is to mimic the structural properties of real world networks which in turn can lead us to models that are able to generate real networks artificially. These models can facilitate experimental and empirical studies for various real world problems.

In this paper, we study a number of network generation models that produce small world-scale free networks. Using a visual analytics method introduced by the authors [38], we show that there are considerable structural differences between networks generated artificially and real world social networks. We also propose a new model to generate artificial social networks with small world and scale free properties.

The rest of the paper is organized as follows: In section 2 we introduce data sets that represent three real world and two hypothetical social networks. In section 3, we discuss a number of different social properties and argue that combining these concepts, we can understand the characteristics required for a real world social network. In section 4, we present a visual analytics method which we use in section 5 to analyze existing artificial models. We then present a network generation model in section 6 which produces small world and scale free networks which are structurally closer to real world social networks. In section 7, we evaluate the proposed model as compared to other models. Next, in section 8, we present comparative results with real world networks to demonstrate the correctness of the proposed model. Finally we conclude and give possible future directions of our research in section 9.

Network	n	e	ad	hd	cc	apl
NetScience	379	914	2.4	34	0.74	6.0
Geometry	3621	9461	2.6	102	0.53	5.31
Imdb	7640	277029	36.26	1271	0.87	2.94

Table 1: n=nodes, e=edges, ad=average degree, hd=highest node degree, cc=clustering coefficient, apl= average path length

2 Data Sets

In order to compare and experiment, we discuss three real world social networks which are described as follows:

NetScience Network is a co-authorship network of scientists working on network theory and experiments, as compiled by M. Newman in May, 2006 [25]. The network was compiled from the bibliographies of two review articles on networks, M. Newman, SIAM Review and S. Boccaletti *et al.*, Physics Reports, with a few additional references added by hand. The biggest connected component is considered for experimentation which contains 379 nodes and 914 edges.

Geometry Network is another collaboration network of authors in the field of computational geometry. The network was produced from the BibTeX bibliography obtained from the Computational Geometry Database 'geombib', version February 2002 (see `http://www.math.utah.edu/~beebe/bibliographies.html`). Problems with different names referring to the same person are manually fixed and the data base is made available by Vladimir Batagelj and Andrej Mrvar: Pajek datasets from the website `http://vlado.fmf.uni-lj.si/pub/networks/data/`. Only the biggest connected component is considered containing 3621 nodes and 9461 edges.

Imdb Network is an actor network where nodes represent actors and two actors are connected to each other if they have acted in a movie together. The data set we use here is a subset taken from the IMDB database (`http://www.imdb.com/`) of movies. This network contains 7640 nodes and 277029 edges.

We also consider two hypothetical networks from our everyday life which are common and easy to comprehend.. Consider a person joining a new organization as an employee and a person joining a sports club as a leisure activity. We refer to these two networks as Employee and Club networks respectively throughout this paper.

3 Structure of Social Networks

In this section, we discuss a number of concepts from the domain of sociology in an attempt to better understand how social networks in the real world are structured.

3.1 Social Ties

People in the real world are linked to each other through social ties. A wide range of ties exist in the society and their study has attracted lots of research activity [35]. The simplest form of a tie is *Dyad* [31] where two people are linked to each other. This is considered as the unit of studying relationships in a social network. *Triads* are relationships between three people and have been the focus of many social network studies [35]. *Groups* of larger size are also possible but since a variety of relationships can form in them, they are less stable [31] and often less studied in sociology. They are often identified by their dense connectivity and clear bounds forming a cluster.

Due to dense interconnectivity, these ties are termed as *strong ties* [19] where nodes that are loosely connected to each other are said to have *weak ties* [11]. A significant work to highlight the importance of these *weak ties* is by Granovetter [11] where he concludes that effective social coordination does not arise from dense interlocking but from the presence of weak ties. Each of us in the society has these weak ties along with strong relationships. These weak ties or acquaintances are important for developing new relationships and possibly joining new social communities. There is a fine mix of both these weak and strong ties that exist in our society and both should be considered to develop a model to generate artificial social networks.

3.2 Homophily

An important human characteristic is *homophily*, tendency of actors or entities to associate with other actors or entities of similar type [28, 29]. Homophily helps to explain why you know the people that you do, because you all have something in common, but one might also wonder how people you know at present determine the people you will know in the future. This also introduces the idea of dynamics in triadic closures. Two people who have a mutual friend will tend to become acquainted in time [28]. A model based on these ideas was proposed by Rapoport who called it *Random Biased Nets*. The idea was to modify the traditional random model of networks such that it incorporates social behaviors. Rapoport also concluded that we occasionally do things that are derived entirely from our intrinsic preferences and characteristics, and these actions may lead us to meet new people who have no connections to our previous friends at all. Although these actions might appear to be random, but can be justified as having strong social background with logical explanations. We limit our study to address this characteristic and refer it as random connectivity pattern. In the light of homophily and social dynamics, we can conclude that new connections between people are formed based on two properties, random connectivity and homophily.

3.3 Extraversion-Introversion

It is interesting to note that in our society, we come across people that are well known and famous, and then there are people who have very few friends and contacts. These ideas are the direct implication of the human trait of extraversion-introversion [17]. Extroverts, who are open to meeting new people and developing new relationships are expected to have high degree of connectivity in a social network as compared to Introverts, who tend to be more reserved, less outgoing, and less sociable.

An important use of this human characteristic is to explain the scale free degree behavior of social networks. A famous person is likely to become more famous as compared to a person who is not well known in the social community.

Termed as the principle of *Preferential Attachment* [1], it explains the growth behavior of networks with power law degree distribution. The idea is that in real world networks, nodes having high degree, have a high probability of attracting more connections as compared to nodes with low connectivity. Thus new social connections have to take this property into consideration as well.

3.4 Observations and Inferences

In our society, we do not form individual relations with people, but with groups of people. These relations are defined by particular circumstances, interests or some context like our school, work place, family [29, 11] and can be explained by homophily. Since these groups are densely connected to each other, often forming cliques, their social ties are considered as strong ties. Since our society is built using these cliques, we call them 'Building Blocks' of our society.

Each of these 'building block' or 'group' is like a small cluster joined to each other by people belonging to more than one group [36]. When these small clusters have many connections to each other, they form bigger size clusters. The size of clusters in a network, vary to a large extent, and so does the number of clusters. Both these parameters depend largely on how the individuals and their ties evolve in a society, how new connections are formed and older ones maintained or destroyed.

Let us consider the example of the actor network. When an actor acts in a movie, the social interactions will take place within the entire cast of the movie and form new ties between actors if they do not exist previously. These interactions will be represented with a clique where all the nodes representing the actors will be connected to each other. The authorship network is no different as people co-authoring an artifact will form a clique. Similarly in the real world, usually groups of larger size are formed. Considering the two hypothetical examples, a new employee will most likely interact with different colleagues in the same organization who work together on the same project or with whom a person shares an office for example. For a person joining a sports club, he will interact with people sharing similar activities instead of just one or two other people. This is to highlight the idea that a person not necessarily interacts with only one or two other people, but more than two people and this is the reason why we obtain cliques of larger sizes in social networks.

Addressing the principle of Preferential Attachment, we argue that for every node in a group (or Clique), few nodes have higher connectivity with other nodes. For example, in a group representing the actors playing in the same movie, the famous actors will have many connections with others as they would have played a role in many movies, and the actors who are starting their career, or are not so well known will have only a few connections. Similarly, in the authorship network, an experienced researcher would have published an artifact with many other researchers and thus would have a high number of connections.

Finally, we look at the society on the whole where we consider the average path length of the networks. One way to have low average path lengths in a network is by random connectivity of nodes, where Watts and Strogatz [37] used this method to have low average path lengths in small world networks.

Combining all these principles, we can conclude that the important elements to capture in the structure of a social network are:

1. Social networks consists of many small groups that are densely connected within themselves forming cliques.
2. These groups overlap due to individuals having multiple affiliations.
3. Some groups have many overlaps which creates large size communities or clusters. to many different groups.
4. A certain degree of randomness exists where we occasionally do things that are derived entirely from our intrinsic preferences and characteristics. These actions lead us to meet new people who have no connections to our previous friends at all.
5. The random connectivity pattern and the presence of high degree nodes is responsible for the low distances between any two people on average.
6. Every group of people has a few Extroverts and many Introverts, where extroverts are responsible for interconnecting people from different domains and the society at large.

We incorporate all these principles in the proposed model. We discuss the details of the proposed model in section 6.

4 Topological Decomposition

In this section, we present a visual analytics method introduced by [38] to analyze networks. The method is based on a decomposition technique which exploits the fact that nodes having high degree are responsible for keeping large size networks as a single connected component. Once, these high degree nodes are removed, the network breaks into smaller components. We visualize these components using graph drawing algorithms. Since the method is based on the topology of the network, we call it 'Topological Decomposition'.

To decompose the network into several components, we use the idea of Max$_d$-Degree Induced Subgraphs (Max$_d$-DIS) [38] where Max$_d$-DIS is an induced subgraph constructed by considering only the nodes having degree at most d in graph G. Mathematically for a graph $G(V,E)$ where V is a set of nodes and E is a set of edges, the Max$_d$-DIS is defined as an induced subgraph $G'(V',E')$ such that $V' \subseteq V$ and $\forall u \in V', deg_G(u) \leq d$ where d can have values from 0 to the maximum node degree possible for the network under consideration.

Consider the example of the Geometry Author network shown in Fig. 1. The graphs are drawn using Fast Multipole Multilevel Method (FM^3) [14] which is a force directed algorithm. These algorithms put nodes densely connected to each other closer in the layout and pushes nodes that are not connected away from each other.

The entire network is shown in Fig. 1(a), where as Fig. 1(b) shows a small portion being focused where the encircled nodes represent densely connected nodes or more precisely cliques. Fig. 1(c) and (d) show portions of the Max$_3$-DIS and Max$_5$-DIS. In these figures, it is quite easy to visually detect the cliques or the densely connected nodes.

We refer back to section 3.4 where we enumerated a number of observations about social networks. The first observation can easily be verified

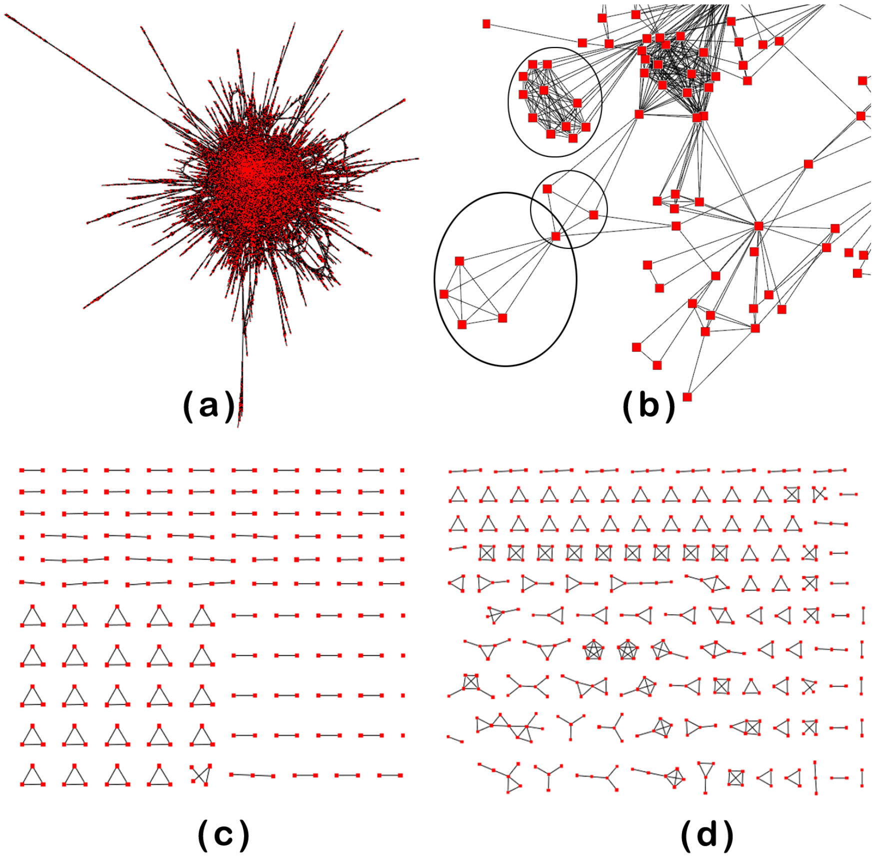


Figure 1: Geometry Co-Authorship Network (a) Entire Network (b) Focus on a Small Portion (c) Part of Max_3-DIS (d) Part of Max_5-DIS

from Fig. 1(c,d) with the presence of cliques in the Geometry network. From Fig. 1(b,d), we can also notice the overlap of cliques to form larger size connected components. As high degree nodes are introduced in this network, the disconnected components start to connect to each other forming one big giant component as shown in Fig. 2(a,b).

The low average path lengths of the three social networks is presented in Table 1. These low values are due to two properties described earlier in section 3.4. These are the random connectivity of nodes which reduces the overall average path length [37] and the presence of very high degree nodes which [38]. Both these can clearly observed when we draw Max_{10}-DIS and Max_{15}-DIS.

5 Related work

In this section, we review a number of network generation models proposed in the literature having small world and scale free properties. A comparative summary of these models is presented in Table 2.

As a general classification, these different models can be grouped into two categories; *Evolving* models and *Static* models. Evolving models are the models that explain the evolution of complex networks as a function of time where the idea is to model the growth behavior of these networks. A good example is the Barabási and Albert model for scale free networks. Nodes are introduced continuously in the network and following the principle of preferential attachment, power-law degree distribution appears. Static models are the models that are concerned with how networks are structured so that certain properties of complex networks are present. Here, the term *structure* means the arrangement of nodes and edges, and how they connect to each other. The Watts and Strogatz model is such an example, as the model starts with a certain number of nodes and edges, that do not increase with the passage of time but explain how high clustering coefficient and low average path lengths appear in a network through rewiring of edges.

Both evolving and static models are of interest as they serve different purposes. Evolving models try to identify the principles that govern the evolution of the physical systems around us. Static models propose methods to understand the structure and formation of networks. Both these types of models can be used to construct artificial networks with properties similar to real world networks to facilitate experimental and empirical studies.

There are two important and common aspects to these networks with a couple of exceptions. The first is that they all use preferential attachment to obtain a scale free behavior for the degree distribution, and the other is that they force the formation of triads in the networks, as a result of which, the clustering coefficient increases as compared to any random network. The triad formation step, results in graphs with high clustering coefficient but cliques of bigger sizes are abscent from these networks as shown in Fig. 4.

Before we review different models to generate small world and scale free networks at the same time, we first present the models to generate only small world and only scale free networks.

Lets have a look at the small world model proposed by Watts and Strogatz. We start with a ring of n vertices in which each vertex is connected to its k nearest neighbors, for a given k. This forms a regular graph as shown in Fig-

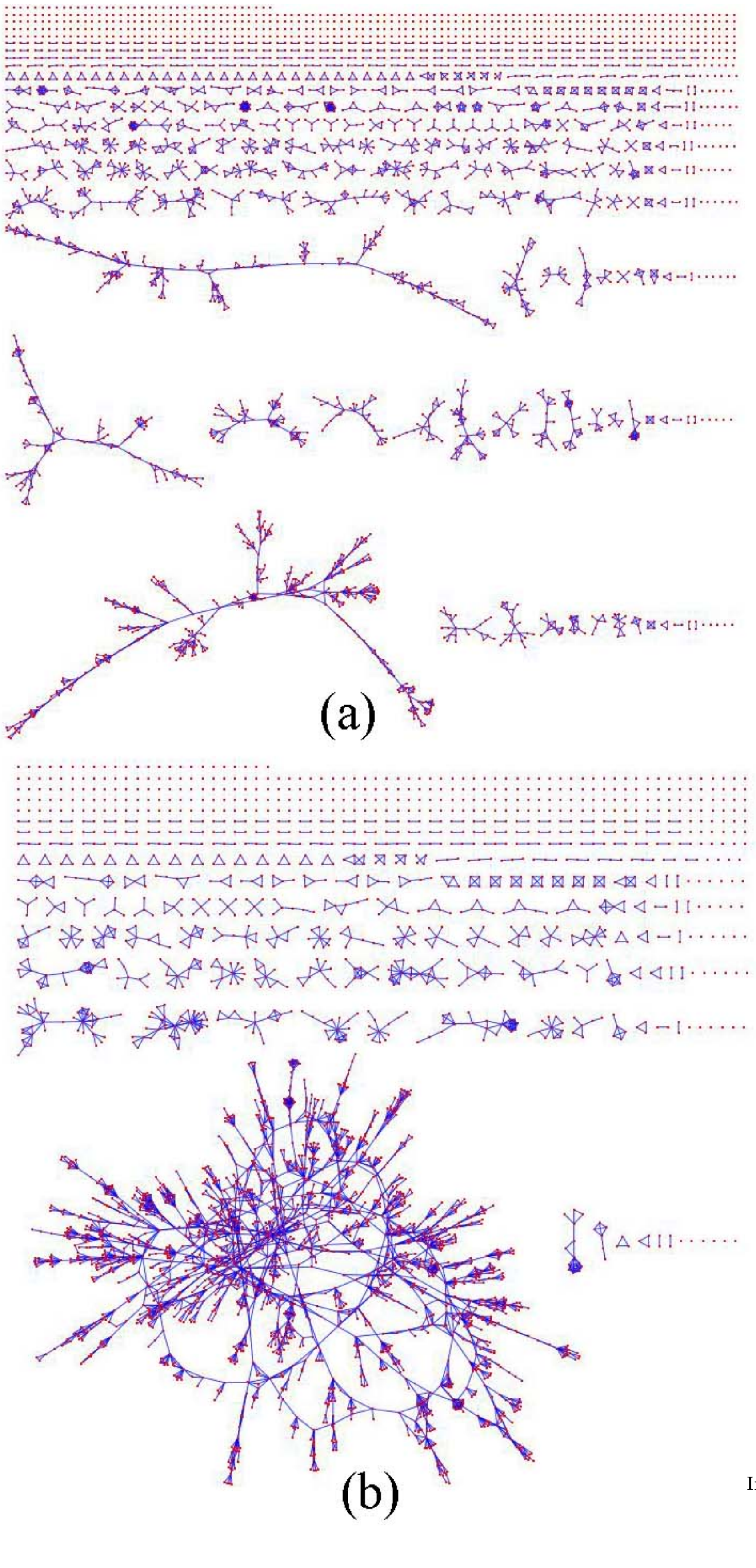


Figure 2: Geometry Co-Authorship Network (a) Max_{10}-DIS (b) Max_{15}-DIS. With the introduction of high degree nodes, small disconnected components start to connect to each other eventually forming one big connected component

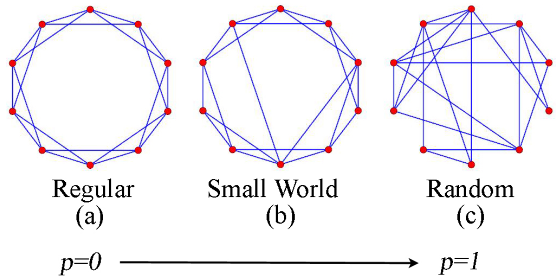


Figure 3: From a Regular network to a Random Network, where random rewiring of few edges in a regular network produces a small world network with high clustering coefficient and low average path length.

ure 3(a). Then, each edge is rewired with a given probability p by choosing randomly a new vertex to connect. In a regular graph, since neighbors are connected to each other, the overall clustering coefficient is very high. On the other hand, the average path length is very low as vertices are only connected to their neighbors. Randomly rewiring a few nodes introduces edges connecting nodes lying at long distances, which in turn, reduces the overall average path length. Since many vertices are connected to their neighbors, the overall clustering coefficient remains high whereas the average path length is reduced, giving us the properties of a small world network (see Figure 3(b)). If the process of random rewiring continues, we eventually end up rewiring every node which results in a random graph as vertices no longer share common neighbors. It is important to note that networks produced using this model do not have scale free degree distribution. Since every vertex in the network initially has a fix k degree, random rewiring of only a few vertices does not effect the overall behavior of the degree distribution. More formal studies of this model have been conducted with interesting results [5].

Barabási and Albert explained how scale free networks emerge in real world networks through another model. To begin, there are n vertices and no edges connecting them. At every time step t, a new vertex v with m edges is added to the network. These edges are connected to existing vertices with the probability proportional to the degree of the nodes in the network. Obviously, at the beginning, when there are no edges, the probability of connection of all the vertices is the same. As the network grows, gradually few nodes begin to have higher node degree and thus higher probability of connecting to newly introduced nodes in the network. This preferential bias in the connectivity is termed as preferential attachment as new nodes prefer to attach to high degree nodes. Mathematical results for scale free graphs have been studied by several researchers such as [2, 3].

Turning towards the models that produce small world and scale free networks, both at the same time.

Holme and Kim [15] modified the well known Barabasi and Albert model [1] to obtain graphs that are small world as well as scale free. The idea is pretty simple and effective. A Triad formation step is added after the preferential attachment step where every node introduced in the network, connects not only to node w, but also to a randomly chosen neighbor of w thus resulting in a triad formation. This results in the formation of lots of triads in the network increasing the overall clustering coefficient. A parameter m_0 is used to decide the initial number of vertices with no edges. Another parameter m is used to decide the number of edges a newly added node will have in the network. This parameter can be used to control the node-edge density of the network. The newly added vertex connects to m different nodes based on the probability which is proportional to their degree. As a result, every new node introduced in the network will form a triad with the highest degree node, which results in lots of triads around high degree nodes. But since the m vertices are chosen solely on the basis of their degree, no clear community structure appears. Another drawback of this model is that it does not generate cliques of larger size as it only forces the presence of triads. We show Max_{15}-DIS of the network generated using this model in Fig. 4. The parameters are set to generate a network of approximately the same size as that of NetScience network.

The idea of Holme and Kim is similar to another model separately proposed by Dorogovtsev *et al.* [6] in the same year where every new node added to the network is connected to both ends of a randomly chosen link where one of the nodes of this link is selected through preferential attachment. Similar behavior is obtained in terms of connectivity as lots of triads are created and the absence of large size cliques remains a drawback. Moreover no other criteria is used to enforce the presence of community structures in the network.

These models inspired Jian-Guo *et al.* to introduce another similar model [21]. The network starts with a triangle and at each time step, a new node is added to the network with two edges. The first edge would choose a node to connect preferentially, and the second edge will choose a node connected to the first node, again based on preferential attachment. This is different from the previous two models where the second node is randomly chosen. No structural changes occur with this modification in terms of cluster formation, the clustering coefficient is increased by the presence of triads but bigger size cliques are still missing and nodes do not attach to each other based on their domain or surroundings but only on their degree.

Wang *et al.* [34] proposed a model to generate random pseudofractal networks with small world-scale free properties. The model starts with two nodes connected through an edge. At each time step, a new node is added with two edges. The new node is connected to the two ends of an edge and the process is repeated for every existing edge in the network. There is obviously no community structure present in the network. We show the evolution of the network in Fig. 5.

Fu and Liao [10] proposed another extension to the Barabasi and Albert model which they called the Relatively Preferential Attachment method. At each time step, the newly introduced node in the network connects to a node w with preferential attachment, the nodes in the immediate neighborhood of w have higher probability of connecting to this new node as compared to other nodes. The only difference in this model with the already proposed models is that the new node can have m edges instead of two edges where the value of m is chosen as an initial parameter which remains constant throughout the execution of the algorithm. As a result, cliques of variable sizes do not appear in the network. For values of m greater than 2, triads appear in the network increasing the overall clustering coefficient.

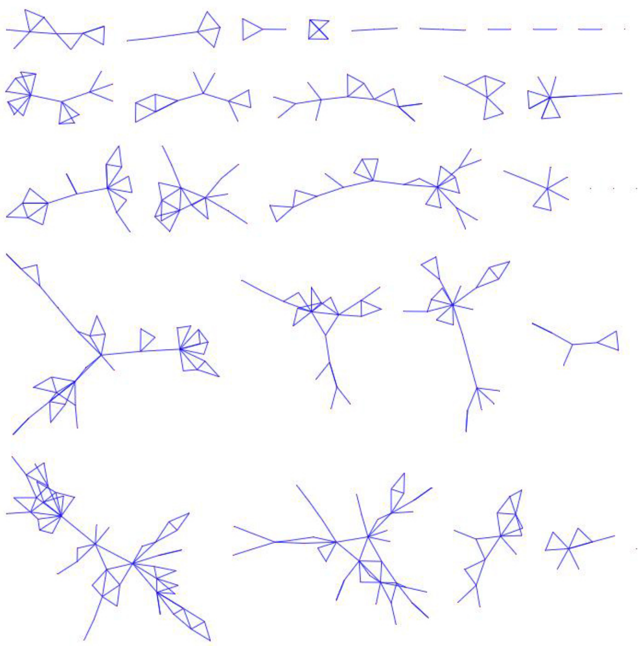


Figure 4: Max$_{15}$-DIS of a network generated using Holme-Kim model with $m_0 = 5$ and $m = 1$, which gives a network of size approximately equal to the NetScience network. Higher degree cliques are clearly missing. The subgraph contains 373 nodes and 584 edges as compared to the entire network with 379 nodes and 757 edges.

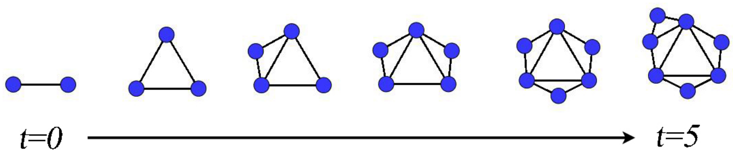


Figure 5: Network generated using Wang *et al.* model for random pseudofractal networks. We see how the network evolves from $t = 0$ to $t = 5$.

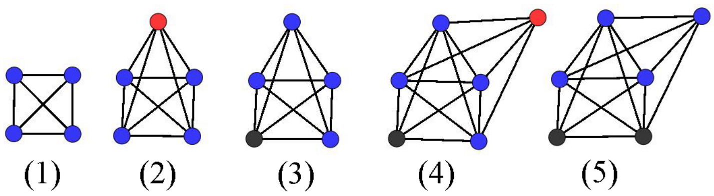


Figure 6: Network generated using Klemm and Eguiluz model.(1) Network starts with $m = 4$ (2) A new node (red) is added connecting to all existing nodes (3) A node is disactivated (black) based on probability proportional to its degree (4) Another node is added (red) (5) Another nodes is disactivated.

Klemm and Eguiluz [18] also proposed a model, where each node of the network is assigned a state variable. A newly generated node is in the *active* state and keeps attaching links until eventually deactivated. At each time step, a new node is added to the network by attaching a link to each of the z active nodes. The new node is set as *active*. One of the existing nodes is deactivated where the probability of a node being deactivated is inversely proportional to its degree i.e lower the degree, higher the probability of deactivation. To reduce the average path length of the entire graph, at every step, for each link of the newly added node, it is decided randomly whether the link connects to the active node or it connects to a random node. Fig. 6 shows the evolution of network and the way new nodes are connected to existing nodes. Again the model does not impose any other constraint so as to form community structures. Fig. 7(a) and (b) show the Max_5-DIS and Max_{10}-DIS of the network generated using this model where the size is approximately equal to that of the NetScience network. We can easily observe that cliques are absent from these subgraphs and the higher clustering coefficient is due to the presence of triads in the network.

Catanzaro *et al.* [4] present a model taking into consideration the assortativity of social networks. At every step, a new node is added to the network based on preferential attachment and a new edge is added between two existing nodes. These existing nodes are chosen on the basis of their degree thus forcing links between similar degree nodes. The model is innovative as it allows addition of new links between old nodes. Since the addition of new nodes is only based on node degrees, nodes of similar degree connect to each other randomly and no clear community structure appears.

Newman *et al.* [26] study models of the structure of social networks with arbitrary degree distributions. The proposed model can also be used to generate networks with scale free degree distribution. The authors introduce the idea to generate affiliation networks similar to co-authorship networks using random bipartite graphs. This idea is used by Guillaume and Latapy [12] as they identify bipartite graph structure as a fundamental model of complex networks by giving real world examples. The two disjoint sets of a bipartite graph are called *bottom* and *top*. At each step, a new *top* node is added and its degree d is sampled from a prescribed distribution. For each of the d edges of the new vertex, either a new *bottom* vertex is added or one is picked among the pre-existing ones using preferential attachment.

A more generalized model based on similar principles was proposed by Bu *et al.* where instead of using the bipartite structure, a network can contain t disjoint sets (instead of just two sets, as is the case of the bipartite graph). In the paper, they discuss the example of sexual web [20] which is based on the bipartite structure. A sexual web is a network where nodes represent men and women having relationships to opposite sex, and similar nodes do not interact with each other. At each time step, a new node and m new edges are added to the network with the sum of the probabilities equal to 1. The preferential attachment rule is followed as the new node links with the existing nodes with a probability proportional to the degree of the nodes.

Wang and Rong [33] proposed a slightly different model, which is still a modified form of the preferential attachment model. Instead of adding one node at a time, the model proposes to add n nodes at each time step which are connected in a ring formation. Any two nodes in the n new nodes are connected to the existing network where these connections are determined through preferential

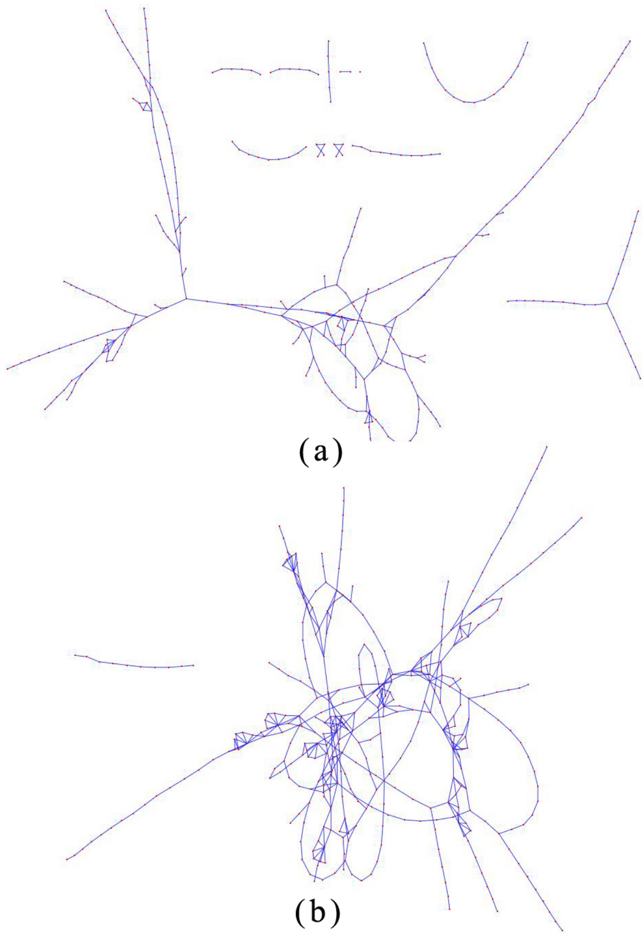


Figure 7: Network generated using Klemm and Eguiluz Model where size is approx. equal to the NetScience network. Figure (a) and (b) show the Max_5-DIS and Max_{10}-DIS respectively. The absence of cliques and the presence of giant component are clearly observable.

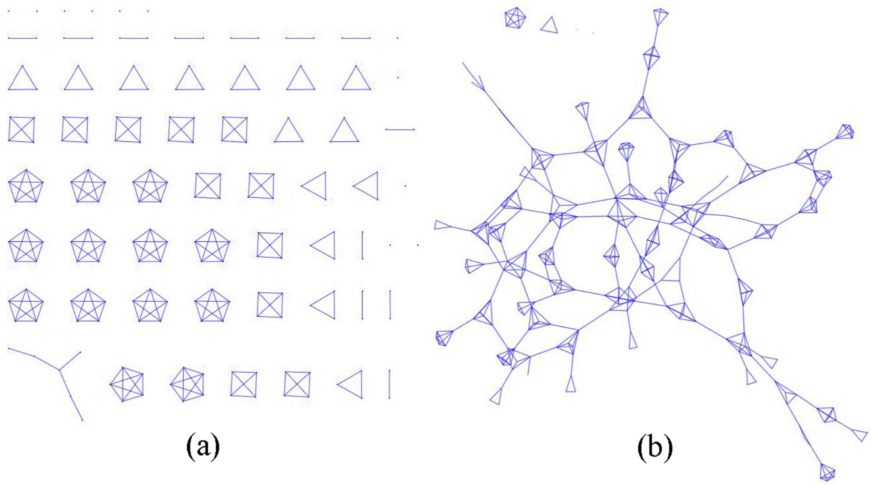


Figure 8: Network generated using Wang and Rong Model where size is approx. equal to the NetScience network. (a) Max_5-DIS: shows the presence of cliques of different sizes (b) Max_{10}-DIS: shows the uniform distribution of these cliques in the network and cliques rarely overlap. Cliques are connected to each other by edges as compared to real social networks where these small social communities overlap to form our society.

attachment. The network breaks into cliques of different sizes but since there is no biased connectivity among the nodes, the cliques are spread uniformly over the network and we cannot find any densely connected set of nodes. Fig. 8(a) and (b) show the the Max_5-DIS and Max_6-DIS respectively. Again the network is generated to be equivalent to the size of the NetScience network. Fig. 8(a) shows the presence of cliques of different sizes and Fig. 8(b) shows the uniform distribution of these cliques in the network without any clear community structure and cliques are connected to each other through edges. As compared to our society, where people belonging to multiple groups connect these small groups forming our connected society at large.

Guo and Kraines [13] proposed a model to study how the clustering coefficient affects the formation of a giant component. The model generates a random social network with finely tunable clustering coefficient. The generator is composed of three steps: first, a degree sequence is generated following a power law. Next, the generator constructs a random network using the algorithm of Molloy and Reed [23]. Finally, the network connections are modified to achieve the desired clustering coefficient. The model is a static one, as it adds all the nodes initially to the network following a prescribed degree distribution. Next, the network is modified to introduce triangles which increases the overall clustering coefficient.

Comparing the different network generation models (See Table 2), the first five models are quite similar to each other, as they try to force the triad formation step, one way or the other. Another common aspect in the first five models is that in every step, only one node and two edges are added to the network. The only other taxonomical grouping possible is the three models where the bipartite and n-partite structures are used as the fundamental property of real world networks. The model of Wang and Rong is slightly different as it allows the addition of m new nodes at every time step. The idea of Klemm and Eguiluz, Catanzaro *et al.* are quite original and provide another way to look at the evolution and structure of complex networks.

6 Proposed Model

As described earlier, the proposed model generates a static network. There are three basic steps in the model which are discussed below.

In the first step, we introduce what we call building blocks in the network. As described in the previous sections, our society is composed of many small groups. So, instead of adding one node at a time, we add cliques of various sizes representing these small groups of the real world. This results in the network having high clustering coefficient. In comparison to various models described earlier, where one node at a time is added to the network. These cliques represent the building blocks of our society as described earlier in section 3.3.

The next step is to join these cliques to form a connected society. These cliques are connected to each other because people belong to multiple groups. From the property of Extraversion-Introversion, we know that there are people with many social contacts as well as people with only a few contacts. These ideas lead us to define for every entity, the number of groups, it belongs to. For a node belonging to two different groups, we simply merge two nodes from different groups, as a result, two cliques are combined with a single node being part of the two cliques as shown in Fig. 10.

To achieve this, we associate a possible connectivity attribute drawn from a degree distribution following power law. Few nodes when being part of many groups, will end up having many social contacts and represent the extroverts in the society.

For each node, this connectivity attribute, called Open connections (OC) determines the number of merges for each node. Note that the number of merges are directly proportional to the final node degree. If a few nodes are merged with many nodes, these nodes will end up with many connections and thus the scale free degree distribution will appear in the network. This attribute is an integer between [1,P] where P is some constant value and represents the maximum node degree a node can have in the network.

Finally, based on these number of merges which represent open connections of nodes (OC), we merge two nodes to build a connected network. We do this by randomly selecting two nodes from the network with open connections (OC). These nodes are merged together. In case, where two nodes of different building blocks are selected and that are already connected to each other by some other node, multiple overlaps appear. This results in small groups connected by more that one node. This represents the phenomena of the real world networks where two small groups are connected to each other by more than two people.

As the network is built from cliques and the connections are directed by scale free degree distribution, we get a network with high clustering coefficient and degree distribution following power law. The average path length of the overall network remains low due to two connectivity patterns, the random connections and the scale free degree distribution. The random connectivity of nodes has been shown to be one of the reasons for low average path lengths by [7, 8].

Comparative Summary of Existing Network Generation Models			
Model, Year	n	m	Innovation
Holme and Kim, 2002	1	m	Triad formation step, forcing a new node to connect to the neighbors of the first node it links to, in order to have triangles and increase the clustering coefficient.
Dorogovtsev *et al.*, 2002	1	2	Randomly chose an edge and attach both ends of this edge with the new node where the probability of choosing an edge is based on the degree of the nodes at its ends.
Jian-Guo *et al.*, 2005	1	2	Each new node attaches to existing node with preferential attachment and choses one of its neighbors again based on preferential attachment (and not randomly as compared to Holme and Kim).
Wang *et al.*, 2006	1	2	For each edge, a new node with two edges is added, which is attached to both end nodes of the edge. Produces Fractals rather than a random graph.
Fu and Liao, 2006	1	m	Once a new node attaches to a node, its neighborhood has a higher probability of connecting to the new node.
Klemm and Eguiluz, 2002	1	m	Activate and deactivate nodes based on node degree where nodes having low degree have a high probability of getting deactivated.
Catanzaro *et al.*, 2004	1	m	Assortativity & Allows growth in old nodes by allowing new edges.
Newman *et al.*, 2002	1	m	Random network following a prescribed degree distribution is generated. Bipartite graphs are used to generate affiliation networks and obtain high clustering coefficient.
Guillaume and Latapy, 2004	1	m	Bipartite Structure identified as a fundamental characteristic for real world graphs (similar to Newman *et al.*, 2002).
Bu *et al.*, 2007	1	m	n-partite Structure, where nodes do not connect to similar node types.
Wang and Rong, 2008	n	m	Add m new nodes and any two nodes in the m new nodes link together from each other and they link to existing nodes based on preferential attachment.
Guo and Kraines, 2009	-	-	Static model that generates a random network with scale free degree distribution for n nodes. Next, the connections are modified to achieve the desired clustering coefficient.

Table 2: Comparing and Summarizing different Artificial Network Generation Models existing in the literature. n=nodes, m=edges

We explain the details of the proposed algorithm below. The following mathematical notations are used throughout the explanation: $G(V, E)$ represents an undirected multigraph where V is a set of n nodes and E is a set of e edges. The graph G is initially empty and the nodes and edges are added as the algorithm progresses. $\mathcal{C}$ represents a set of cliques such that $\mathcal{C} = \{C_1, C_2, \cdots, C_k\}$ are different cliques each comprising of several nodes.

6.1 Step 1: Building Blocks

In contrast to existing network generation models, instead of adding one node or triad at a time, to generate the network, we start by adding cliques of variable sizes to G. Recall from section 3.4, we identified cliques as one of the fundamental patterns present in networks and the Author and Actor network considered as examples here have cliques by construction.

The algorithm takes as parameter, the number of cliques to be generated (k), the minimum (minSize) and the maximum size (maxSize) of the cliques to be generated. A random number is generated between these two limits and for each random number, a clique C_i is added to the graph G such that nodes and edges of the clique become members of V and E respectively. As a result, G contains nodes that are well connected to each other as a clique, and nodes from different cliques are not connected to each other. G becomes a graph comprising of $\mathcal{C} = \{C_1, C_2, \cdots, C_k\}$ as shown in Fig. 9.

If we use a random number generator, for large values of k, the distribution will be equally spread and we will have the same number of cliques for all possible size values. In real networks, this might not be the case as often, cliques of large sizes are rare compared to cliques of small sizes. To take the correct decision, it is important to understand what type of network we are trying to generate. If the network to be generated is expected to have cliques of varying sizes equally distributed, the random generation will serve well our purpose. On the other hand, if we expect that all the cliques will have the exact same size, the minSize and maxSize parameters can be set to that exact value to have all the cliques of the exact same size. And in the case where we expect a non-uniform distribution of different sizes, we can draw the different sizes of cliques using the type of distribution we require our final network to follow. The parameters minSize and maxSize can also be used to control the node edge density. If the values of these parameters are set as 1 and 5 respectively, the cliques generated will have nodes of degree 0 and 1, which in turn, will reduce the overall node edge density. On the other hand, if we want to increase the node/edge density, we can set high values of minSize and maxSize which will generate dense group of nodes and increase the overall node/edge density.

The real networks that we are using for analysis do not have a uniform distribution of cliques. Since these real networks contain many nodes with degree between 1 and 4. While generating networks of equivalent size, we take this information into account and ensure the increased presence of these small degree cliques. For every iteration, whenever a random number is generated having a low degree, another one is added of the same size. Thus for every random number generated between 1 and 4, we add two cliques instead of one. Experimental results show that this method is effective as we get networks similar to real world networks.

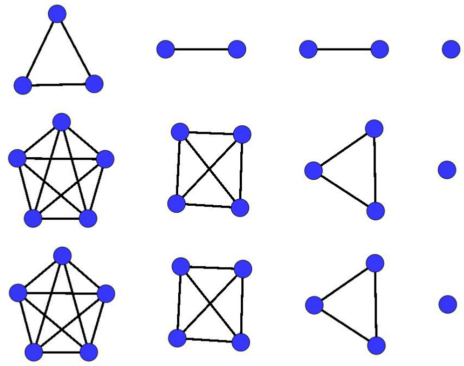


Figure 9: Step 1: Network after execution of step 1 with minSize=1, maxSize=5 and k=10.

We consider the example of co-authorship network and explain how these values effect the algorithm. We use $k = 10$, minSize=1 and maxSize=5 and a random generation for the size of the cliques. After the execution of this step, we get a network as shown in Fig. 9. The idea of introducing cliques, comes from the work of [26, 12] where affiliation networks and the bipartite structure was identified as an important structural property of the way, the Author and the Actor networks are constructed in the real world. People interact to co-author an artifact, as a result we get cliques representing an artifact. The idea is equally applicable to the Actor network, where the cast of every movie forms a clique. This phenomena was explained in detail in section 3 earlier and equally holds for the Employee and Club network.

Note that the size of the cliques can be forced to be exactly 3, in which case we would have forced the presence of only triads just as the other network generation models presented in section 5. Due to the presence of cliques (or triads), the average clustering coefficient of the entire graph increases as compared to a random graph which is a fundamental property to identify a small world network.

6.2 Step 2: Determine Number of Merges

Since our goal is to control the frequencies of the node degrees, we want to enforce a certain degree distribution. In order to have the degree distribution of G follow a scale free behavior, we generate a scale free degree distribution using a power law function. We associate this distribution on the nodes of graph G as an attribute and call this as open connections OC. This attribute is used to determine how the nodes are interconnected to each other in the next step.

An important variation to this step can be the assignment of an equal value to all nodes. As a result, the network produced will have only small world properties, i.e. high clustering coefficient and small average path length. The equal value assignment will ensure that the degree of all the nodes is approximately equal and thus the final degree distribution will not follow a power law, rather a Poisson distribution.

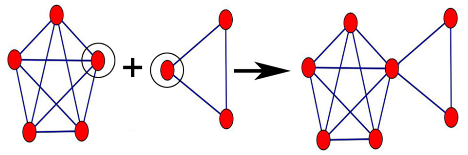


Figure 10: Merging two nodes from two different cliques so that a node becomes part of two cliques.

6.3 Step 3: Merge Nodes

Two cliques can be combined by considering that one or more than one common authors are part of the two cliques, and these nodes play the role of combining these cliques (see Fig. 10). This is true for other real world networks as discussed earlier in section 3.

Merging two nodes creates connections between previously disconnected cliques. Moreover, the merged node plays the role of a bridge between the two small clusters. In terms of the degree, the node gets many new connections. The more the node is merged with other nodes, the more it gets connections and higher would be its node degree. This is the reason why we draw the number of merges from a power law function, as a result, the final degree distribution follows a power law.

An important decision while merging two nodes say n_1 and n_2 with OC values oc_1 and oc_2 is, how to decide the oc_n for the new node n_n. We experimented with the following different methods:

- Max: Assign the new node the maximum of the two OC values $oc_n = Max(oc_1, oc_2)$
- Min: Assign the new node the minimum of the two OC values $oc_n = Min(oc_1, oc_2)$
- Avg: Assign the new node the average of the two OC values $oc_n = Avg(oc_1, oc_2)$
- Rand: Assign the new node one of the two OC values randomly $oc_n = Rand(oc_1, oc_2)$

Assigning maximum value forces the degree distribution of the network to take a more linear decay as most of the low degree nodes disappear quickly from the network and lots of high degree nodes are left for connectivity. On the other hand, assigning minimum value removes the few nodes with very high degree and the characteristic long tail in the degree distribution disappears from the network. As similar behavior is observed with the average assignment as the long tail disappears and the average node degree increases with this assignment. The best results are obtained by a random assignment as nodes with high and low degree are equally removed and thus the overall degree distribution follows scale free behavior. We show the experimental results using the random method in section 8.

7 Evaluating Generated Networks

The proposed model is very close to the model proposed by Guillaume and Latapy [12] or that of Newman [26]. Although our approach is slightly different from these two models. We differentiate between the connectivity within group and connectivity in the society. The connectivity within group depends on the building blocks, which in this case are cliques. Connectivity in the society depends on the human trait of Extraversion and Introversion. The connectivity within group is responsible for the high clustering coefficient, as opposed to many other models where forcing triads raises the over all clustering coefficient. The connectivity with the society is responsible for the overall degree distribution following power-law. These steps can be modified in the model to obtain networks with different properties. For example, if we modify the number of merges drawn from the scale free behavior to follow a Poisson distribution, the model will produce networks which are only small world and not scale free. On the other hand, if we modify the building blocks by replacing the cliques by a star-like structure, where one node is connected to many nodes, we will get networks with only scale free properties with low clustering coefficient.

Thus, as compared to the model of Guillaume-Latapy [12] and Newman [26], from the proposed model, we are able to capture the principles of the bi-partite structure of many real world networks by introducing a different approach. Moreover the same model can be used to generate scale free networks with a simple modification to the network. We leave the proof of this variation as part of future work.

Next, we evaluate the networks generated by the proposed model using the Max_d-DIS decomposition. Fig. 11(a) shows the entire network generated where the network has size similar to NetScience network. Fig. 11(b) shows the Max_5-DIS of the network where the network breaks into small connected components just as the co-author networks in Fig. 2 and Fig. 1.

Similar observations can be made about the network generated equivalent in size to the Geometry network. Fig. 12(a) shows the Max_5-DIS of the network where the network breaks into small connected components just as the geometry network studied in section 4. Fig. 12(b) shows the Max_{10}-DIS of the network with the appearance of the giant component. Since the model is based on cliques as the building blocks to construct the entire network, it is obvious that using the topological decomposition, we find the presence of these small densely connected group of nodes. As the node degree is increased, in the case of Fig. 12(b) where Max_{10}-DIS contains nodes of at most degree equal to 10, we find a similar behavior in the connectivity of nodes just as we analyzed in section4, a phase shift take place and a single giant [16, 9] connected component appears.

Fig. 13 shows the degree distribution of the networks generated using the proposed model. We have generated networks of size equivalent to three social networks, the NetScience, Geometry and Imdb network. The degree distribution clearly shows that the networks generated follow the power law.

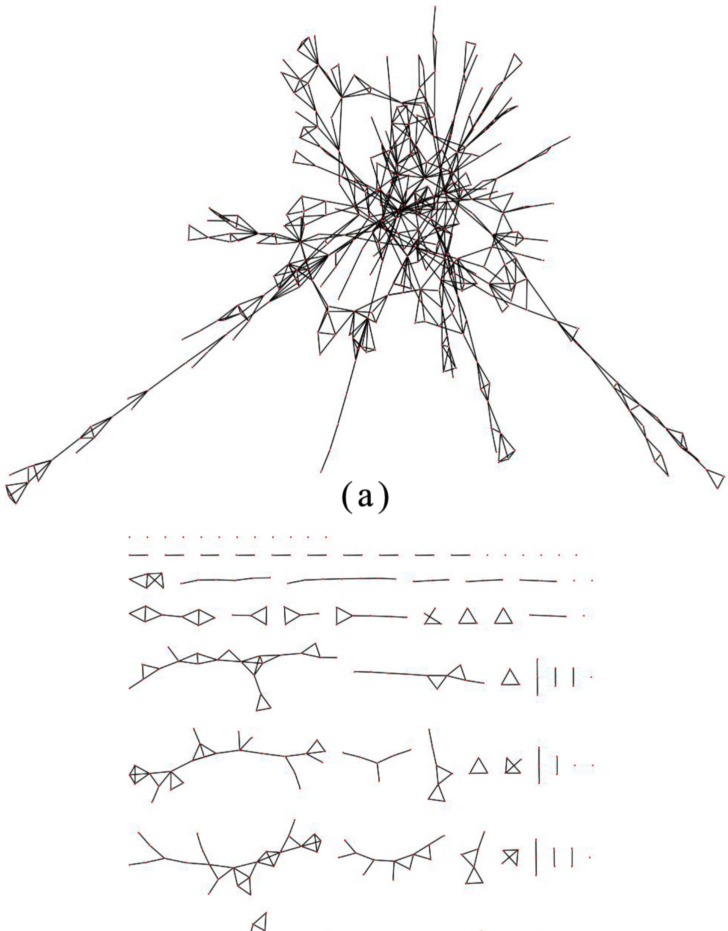
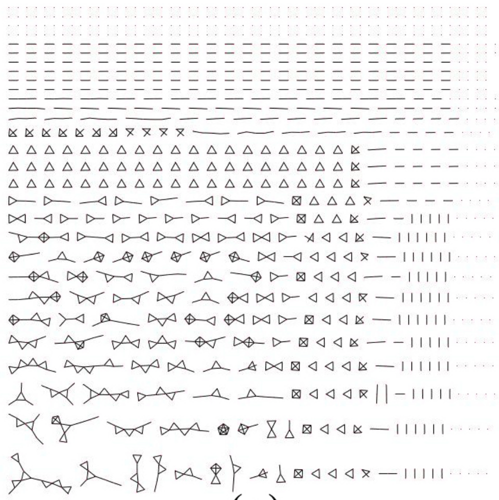
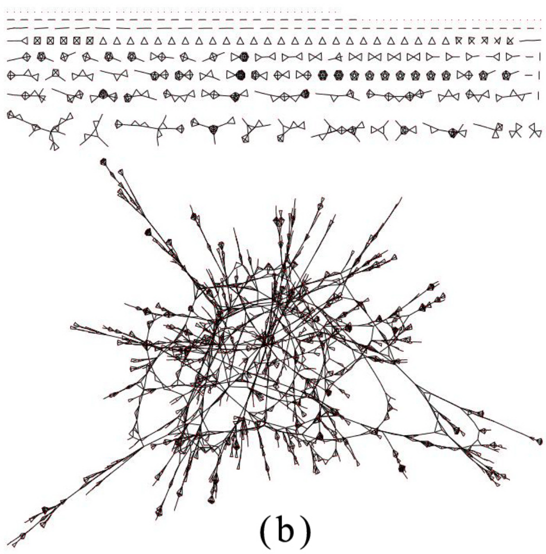


Figure 11: Network generated using proposed network model where the size is approx. equal to NetScience network. cliques=200 minSize=1, maxSize=7 (a) Entire network (b) Max_5-DIS.



(a)



(b)

Figure 12: Network generated using proposed network model where the size is approx. equal to Geometry network. cliques=3000 minSize=1, maxSize=9 (a) Max_5-DIS (b) Max_{10}-DIS.

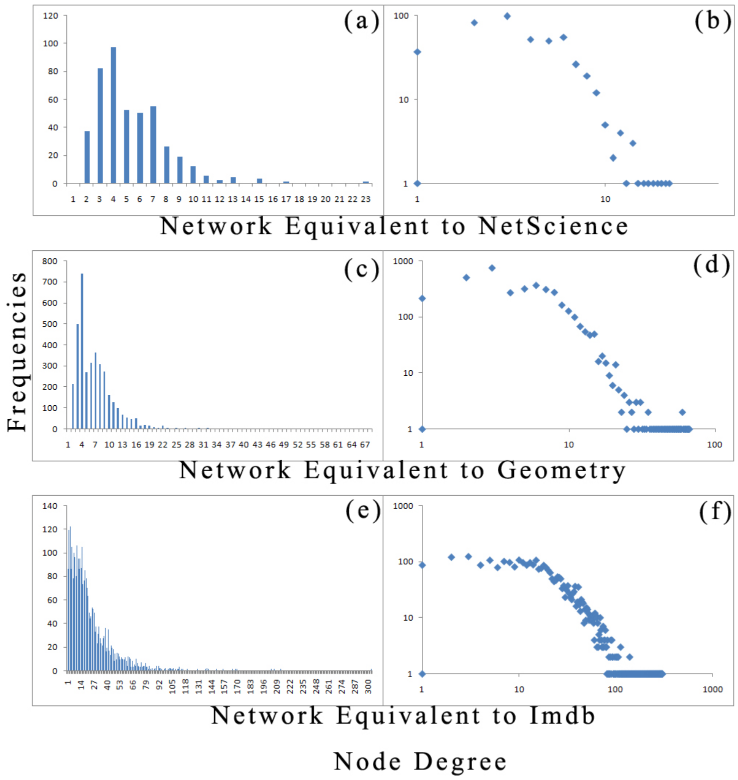


Figure 13: Degree Distribution of equivalent size networks generated using the proposed Model. (a,c,e) Represent the bar charts and (b,d,f) represent the Log-Log plot of the Frequency-Degree distribution.

Comparison between NetScience and Other Network Models					
Model	Nodes	Edges	APL	CC	HD
NetScience	379	914	6.04	0.74	34
Random Graph	379	914	3.94	0.01	11
Zaidi *et al.*	364	935	4.7	0.65	22
Holme and Kim	379	757	4.86	0.77	42
Fu and Liao	379	744	4.03	0.75	31
Klemm and Eguiluz	379	755	6.40	0.5	24
Catanzaro *et al.*	379	898	2.42	0.58	197
Guillaume & Latapy	379	5315	2.30	0.54	109
Bu *et al.*	379	755	3.05	0.37	80
Wang and Rong	379	943	4.32	0.37	14

Table 3: Comparing different models with the Collaboration Network of Scientists from the NetScience data. APL=Average Path length, CC=Clustering Coefficient, HD=Highest Node Degree

Comparison between Geometry and Other Network Models					
Model	Nodes	Edges	APL	CC	HD
Geometry	3621	9461	5.31	0.53	102
Random Graph	3621	9461	5.15	0.001	15
Zaidi *et al.*	3682	10928	5.71	0.65	67
Holme and Kim	3621	7241	7.3	0.79	90
Fu and Liao	3621	10662	4.22	0.72	101
Klemm and Eguiluz	3621	10857	2.27	0.72	197
Catanzaro *et al.*	3621	8896	2.47	0.48	1720
Guillaume & Latapy	3621	528499	*	*	1275
Bu *et al.*	3621	10856	3.13	0.24	607
Wang and Rong	3621	10828	4.6	0.10	30

Table 4: Comparing different models with the Collaboration Network of Scientists from the Computational Geometry data. APL=Average Path length, CC=Clustering Coefficient, HD=Highest Node Degree

Comparison between Actor and Other Network Models					
Model	Nodes	Edges	APL	CC	HD
Imdb	7640	277029	2.94	0.87	1271
Random Graph	7640	277029	2.48	0.009	102
Zaidi *et al.*	7413	244905	3.1	0.98	352
Holme and Kim	7640	274865	2.35	0.09	2303
Fu and Liao	7640	29972	4.00	0.76	163
Klemm and Eguiluz	7640	274374	1.99	0.97	7627
Catanzaro *et al.*	7640	28127	1.99	0.78	7639
Guillaume & Latapy	7640	2378281	*	*	2614
Bu *et al.*	7640	274935	1.99	0.83	12151
Wang and Rong	7640	273355	3.28	0.94	83

Table 5: Comparing different models with the Imdb network from the IMDB dataset. APL=Average Path length, CC=Clustering Coefficient, HD=Highest Node Degree

8 Results and Discussion

We have used the NetScience, Geometry and Imdb data sets for a comparative study. These are well studied examples of social networks and have been used by several researchers for empirical and experimental studies.

We calculate a number of statistics using various Network generation models and compare them with the real world networks of equal sizes. The results are shown in Table 3, Table 4 and Table 5. We have included the statistics for a random network for the three data sets. In some cases, the models are not parameterized and thus the node-edge density could not be controlled. We tried to generate models of similar size in terms of number of nodes, and where possible, similar number of edges. An important observation about these networks is that since all of them use the preferential attachment to produce the scale free property, the degree distribution for all the models follow a power law. To the best of our knowledge, there is no metric which tries to identify the presence of communities in a network by analyzing the graph on the whole in a global perspective, thus the presence of community structure in the proposed model is only justified by construction.

Looking at some individual results for the various models in comparison to the real world networks. For example, graphs generated using the model of Guillaume and Latapy, the node-edge density in every case is very high and could not be controlled. The model of Fu and Liao, in all the three examples, have a very low clustering coefficient as compared to the respective real world network and thus could not really be classified as generating similar networks to the real world networks used as examples in our study. Looking at the clustering coefficient of the model by Wang and Rong in Table 4, it is quite clear that the model fails to generate a high clustering coefficient for a similar size network. An observation about the model of Holme and Kim, In Table 5, where the node-edge density of the network is comparatively high to other two networks but the the network has a large size, the clustering coefficient drops considerably. The model of Klemm and Eguiluz scales well in terms of clustering coefficient, and the average path length can controlled through a parameter (see Table 3) which gives a good approximate result. Also, from Table 5, the average path length in case of a number of models is 1.99, which is a direct implication of a node having a very high degree. As a result, most of the nodes are connected to this high degree node and thus have a low average path length of the entire network.

From the above examples, one obvious problem that can be inferred is that these models have problems with scalability, as the node edge density is varied for a network, the models are not able to reproduce comparative values with real world networks for various statistics. On the other hand, the proposed model in this paper has the ability to control the size of cliques as the starting point, which helps us to control the density and at the same time, generate small world and scale free networks. The values are quite close to the ones expected and thus the proposed model is quite flexible.

9 Conclusion and Future Research

In this paper, we have studied the concepts of homophily, triads and preferential attachment as important properties for the structure of social networks. We use these concepts to present a model to generate artificial social networks. We evaluated a number of network generation models that successfully generated small world and scale free networks but produced structurally different networks as compared to real world networks. Results show that the proposed model indeed generates networks that are topologically similar to real world networks as compared to the other existing models.

We intend to extend our study to other types of networks such as biological and technology networks. Although these networks have small world and scale free properties but they are again structurally different from social networks and thus we need to modify the proposed model to mimic the behavior of these other types of networks.

References

[1] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[2] B. Bollobás. Mathematical results on scale-free random graphs. In *In Handbook of Graphs and Networks*, pages 1–34. Wiley-VCH, 2003.

[3] E. M. Bollt and D. ben Avraham. What is special about diffusion on scale-free nets? *New J. Phys.*, 7:26, 2004.

[4] M. Catanzaro, G. Caldarelli, and L. Pietronero. Assortative model for social networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 70(3), 2004.

[5] S. Dorogovtsev and J. Mendes. Exactly solvable small-world network. *European Physics Letters*, 50(1):1–7, 2000.

[6] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advances in Physics*, 51:1079–1187, June 2002.

[7] P. Erdös and A. Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[8] P. Erdos and A. Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[9] J. W. Essam. Percolation theory. *Reports on Progress in Physics*, 43(7):833, 1980.

[10] P. Fu and K. Liao. An evolving scale-free network with large clustering coefficient. In *ICARCV*, pages 1–4. IEEE, 2006.

[11] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, May 1973.

[12] J.-L. Guillaume and M. Latapy. Bipartite graphs as models of complex networks. In *Workshop on Combinatorial and Algorithmic Aspects of Networking (CAAN), LNCS*, volume 1, 2004.

[13] W. Guo and S. B. Kraines. A random network generator with finely tunable clustering coefficient for small-world social networks. In *CASON '09: Proceedings of the 2009 International Conference on Computational Aspects of Social Networks*, pages 10–17, Washington, DC, USA, 2009. IEEE Computer Society.

[14] S. Hachul and M. Jünger. Drawing large graphs with a potential-field-based multilevel algorithm. *Graph Drawing*, pages 285–295, 2005.

[15] P. Holme and B. J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65:026107, 2002.

[16] S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *ArXiv Mathematics e-prints*, Sept. 1993.

[17] C. J. Jung. *Psychologischen Typen*, volume Translation H.G. Baynes, 1923. Rascher Verlag, Zurich, 1921.

[18] K. Klemm and V. M. Eguiluz. Growing scale-free networks with small world behavior. *Physical Review E*, 65:057102, 2002.

[19] D. Krackhardt. The strength of strong ties: the importance of philos in networks and organization. In N. Nohria and R. G. Eccles, editors, *Networks and Organizations*. 1992.

[20] F. Lilijeros, C. Edling, L. Amaral, E. Stanley, and Y. åberg. The web of human sexual contacts. *Nature*, 411:907–908, 2001.

[21] J.-G. Liu, Y.-Z. Dang, and Z. tuo Wang. Multistage random growing small-world networks with power-law degree distribution. *Chinese Phys. Lett.*, 23(3):746, Oct. 31 2005. Comment: 3 figures, 4 pages.

[22] S. Milgram. The small world problem. *Psychology Today*, 1:61–67, May 1967.

[23] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–180, 1995.

[24] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167, 2003.

[25] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 74(3), 2006.

[26] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 1):2566–2572, February 2002.

[27] N. Päivinen. *A Quest for the Hidden Knowledge.* PhD thesis, University of Kuopio,Finland, 2007.

[28] A. Rapoport. Contribution to the theory of random and biased nets. *Bulletin of Mathematical Biophysics*, 19:257–277, 1957.

[29] A. Rapoport and W. J. Horvath. A study of a large sociogram. *Behavioral Science*, 6(4):279–291, 1961.

[30] J. P. Scott. *Social Network Analysis: A Handbook*. SAGE Publications, January 2000.

[31] G. Simmel and K. H. Wolff. *The sociology of Georg Simmel / translated and edited with an introduction by Kurt H. Wolff.* Free Press, Glencoe Ill, 1950.

[32] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969.

[33] J. Wang and L. Rong. Evolving small-world networks based on the modified ba model. *Computer Science and Information Technology, International Conference on*, 0:143–146, 2008.

[34] L. Wang, F. Du, H. P. Dai, and Y. X. Sun. Random pseudofractal scale-free networks with small-world effect. *The European Physical Journal B - Condensed Matter and Complex Systems*, 53:361–366, 2006.

[35] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications.* Cambridge University Press, Cambridge, 1994.

[36] D. J. Watts. *Six Degrees: The Science of a Connected Age*. W. W. Norton & Company, 1st edition, February 2003.

[37] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.

[38] F. Zaidi. *Analysis, Structure and Organization of Complex Networks.* PhD thesis, Ecole doctorale de Mathématiques et Informatique, Université de Bordeaux, France., 2010.

Contents

**RESEARCH CENTRE
BORDEAUX – SUD-OUEST**

351, Cours de la Libération
Bâtiment A 29
33405 Talence Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399